Accuracy Paradox. "If you don't know anything about… | by Tejumade Afonja

🔘 **towardsdatascience.com**/accuracy-paradox-897a69e2dd9b

What is Accuracy Paradox?

Accuracy is defined as the freedom from mistake or error. For example, a piece of information is accurate if it represent what is being discussed.

Paradox is a statement that is seemingly contradictory or opposed to and yet is perhaps true.

> Have you considered that the phrase "ignore all the rules" … is a rule itself — uhm! a paradox

In Machine Learning lingo, Accuracy is the proportion of **correctness** in a classification system.

Meaning, if we have a Spam detection system, and out of 5 mails we received, 4 were classified by a Model X as Spam and these mails were indeed Spam. We would say Model X has 80% Accuracy i.e you can rely on Model X, 80% of the time.

```
Accuracy of Model X = 4/5 * 100 = 80%
```

So, What exactly is an Accuracy Paradox? We'll get to that later but first let's consider a case study as a way of an example.

Case Study

Hawkins hospital software team (Will, Dustin, Mike and Lucas) built a classification model for diagnosing Breast Cancer in women. A sample of 1000 women were studied in a given population and 100 of them with Breast Cancer while the remaining 900 were without it. Hawkins software team trained their model based of this dataset. They split the dataset into 70/30 train/test set.

The accuracy was excellent and they deployed the model.

Alas! a couple of months after deployment, some of the women who were diagnosed by the hospital as having "no breast cancer" started showing symptoms of Breast Cancer.

How could this be?

This raised a series of questions and fear amongst the entire population. Hawkins hospital had to do something about this as more and more patient started showing symptoms of Breast Cancer.

They decided to hire a Machine Learning Expert, Bob to help them understand what their software team got wrong considering the fact that the model had an **accuracy** of about 90%.

Hawkins Model's Overview

By splitting the data set, we have

Training set : No breast cancer = 70/100 × 900 = 630
Breast cancer = 70/100 * 100 = 70

Test set : No breast cancer = 30/100 × 900 = 270
Breast cancer = 30/100 * 100 = 30

Say, **we have an assumption H,** that Joyce is suffering from Breast Pain and not Breast Cancer but **another assumption Ho**, (that counters the first) says Joyce is suffering from Breast Cancer.

is true (positive) — Breast Cancer

is false (negative) — No Breast Cancer

The table below represents what happens if this **other assumption Ho** is true or not.

	Null Hypothesis(Ho) is **valid**: Breast Cancer	Null Hypothesis(Ho) is **invalid**: No Breast Cancer
Accept Ho: I think Joyce has Breast Cancer	**TP–** Sad News, Joyce indeed has Breast Cancer	**FP–** False Alarm, Joyce just have a temporary breast pain, no Biggie, no Breast Cancer but of course, Joyce will probably live her whole life thinking she'd die soon.
Reject Ho: I don't think Joyce has Breast Cancer	**FN–** Deadly! Joyce thinks she's healthy when in fact she's slowly dying of Breast Cancer.	**TN–** Good News, Joyce indeed doesn't have Breast cancer

This is called confusion matrix — i hope it's not confusing.

Where:

- TP = True Positive
- FP = False Positive
- FN = False Negative

- TN = True Negative

Hawkins Model's Prediction Result

After training their model with 70% of the data set, Hawkins scientist tested this model with the remaining 30% data to evaluate the model for its accuracy. Their model got 270 prediction right out of the 300.

```
Hawkins Accuracy = 270 / 300 = 0.9
```

This looked like a pretty convincing model with an accuracy of 90%, why then did it fail?

Solving The Mystery

Bob re-evaluated their model and below is the breakdown:

```
Number of women  breast cancer and classified as breast cancer (FN) = 30Number of
women with breast cancer and classified as breast cancer(TP) = 0Number of women
without breast cancer and classified as  breast cancer (TN) = 270Number of women
breast cancer and classified as breast cancer (FP) =
```

Bob represented this with a confusion matrix table below.

	Actual – Cancer	Actual – NOT Cancer	Total
Predicted – Cancer	TP = 0	FP = 0	0
Predicted – NOT Cancer	FN = 30	TN = 270	300
Total	30	270	300

confusion matrix for Hawkins Model

In summary,

Hawkins model **correctly** classified 270 women who do not have breast cancer as **"NO Breast Cancer"** while it **incorrectly** classified 30 women who have breast cancer as **"NO Breast Cancer".**

> We gave Hawkins Model 300 questions and it got 270 answers right. The model scored 270/300 — so, we'd say the model passed excellently right? but did it?

Bob noticed that the model has conveniently classified all the test data as **"NO Breast Cancer"**.

Bob calculated the accuracy of this model which is correct 90% all the time.

```
Accuracy = (TP + TN) / (TP+TN+FP+FN)Accuracy = (0 + 270) / (0 + 270 + 0 + 30)=
0.90Accuracy in % = 90%
```

Bob noticed a pattern, **none** of the "Breast Cancer" data was correctly labeled. Bob went a little further as an expert, he decided to see how the model is doing in terms of **Precision** and **Recall**.

Bob thought to himself, if **none** of these people who have Breast Cancer came out as "having Breast Cancer" this model isn't a **precise** model and neither is it going to **recall** anything except "NO Breast Cancer".

He proved this using the formula for Precision and Recall below:-

```
Precision = TP / (TP + FP)Precision = 0 / (0 + 0) = 0Precision in % = 0%Recall =
TP / (TP + FN)Recall = 0 / (0 + 30) = 0Recall in % = 0%
```

What this means is that, this model will always classify any data passed into it as "NO BREAST CANCER".

This explains why some of the patients were showing symptoms of Breast Cancer.

Hawkins Model simply did not work.

> Bob cracked it, Hawkins Model is a Demogorgon (scam).

Bob proposed a slightly modified Model, he wanted them to understand what he meant by their Model being a Demogorgon.

Bob's Model

After training his model with 70% of the data set, Bob then tested the model with the remaining 30% data to evaluate the model. Bob, unlike Hawkins software team did not rely on accuracy as the only metric to evaluate his model.

Below is the result of Bob's Model;

```
Number of women  breast cancer and classified as breast cancer (FN) = 10Number of
women with breast cancer and classified as breast cancer(TP) = 20Number of women
without breast cancer and classified as  breast cancer (TN) = 200Number of women
breast cancer and classified as breast cancer (FP) = 70
```

	Actual – Cancer	Actual – NOT Cancer	Total
Predicted – Cancer	TP = 20	FP = 70	90
Predicted – NOT Cancer	FN = 10	TN = 200	210
Total	30	270	300

confusion matrix for Bob's Model

Clearly, Bob's model made some mistakes by scaring 70 perfectly healthy people who don't have Breast Cancer. **Bob thought to himself, isn't it better to think you have Breast Cancer and not have it than to think you don't have Breast Cancer but you've got it.

Bob's Model Evaluation

Accuracy = (TP + TN) / (TP + TN + FP + FN)Accuracy = (20 + 200) / (20 + 200 + 10 + 70) = 0.733Accuracy in % = 73.3%

Questions arose within the software team of Hawkins, how could Bob possibly tell them that his model (with 73% accuracy) is better than theirs (with an accuracy of 90%.)

Bob went further to calculate Precision and Recall of his new model

Precision = TP / (TP + FP)Precision = 20 / (20 + 70) = 0.222Precision in % = 22.2%Recall = TP / (TP + FN)Recall = 20/ (20 + 10) = 0.67Recall in % = 67%

Although Bob's model assumed 90 women instead of 30 had cancer in total, it predicted Breast Cancer correctly 22.2% of the time as opposed to Hawkins Model with Precision of 0.

Also, out of the 30 women that actually has Breast Cancer, Bob's Model was able to correctly recall that someone has Breast Cancer 67% of the time as opposed to Hawkins Model which has 0 recall.

After this, Bob was able to convince the team that his model was better than what they currently have.

But what about the difference in Accuracy, Dustin asked?

Bob replied: it's a Paradox.

Accuracy Paradox

Accuracy Paradox for predictive analytics states that predictive models with a given level of accuracy may have greater predicting power than models with higher accuracy

Breaking this down,

Predictive models with a given level of accuracy (73% — Bob's Model) may have greater predictive power (higher Precision and Recall) than models with higher accuracy (90% — Hawkins Model)

And that's why it's called a Paradox because, intuitively, you'd expect a Model with a higher Accuracy to have been the best Model but Accuracy Paradox tells us that this, sometimes, isn't the case.

So, for some systems (like Hawkins') Precision and Recall are better than the "Good ol' Accuracy" and thus it's important to use the appropriate metrics to evaluate your model.

Bob the brain, saved the day!

*** it's important to note that trading False Positives could come at a very deadly cost as*

While a search engine model could probably make do with a few false positives, a breast cancer model shouldn't as this could lead to perfectly healthy people being introduced to a brutal treatment or worse.